Venmo Data Project Report

Yifan Zhu

Mario Ma

Jenny Deng

Vicky Chen

August 10, 2024

Venmo is a peer-to-peer (P2P) mobile payment app owned by PayPal. The Venmo app allows its users to exchange money with just a click of a button. In the fourth quarter of 2019, Venmo's net payment volume amounted to 29 billion U.S. dollars, representing a 56 percent year-on-year growth, and its user-base had more than 40 million active accounts. What made Venmo so popular in the US is its social flavor; users are required to accompany their transactions with a message describing what the transaction was about. This social twist has allowed Venmo to transform financial transactions into sharing experience.

In this project, we are given a sample of Venmo's transaction dataset. Due to the large sample size, the entire project was coded in Spark. Our goal is to have a comprehensive understanding of user's spending behavior profile, user's social network, and predict in advance how many times you customers will transact. To answer these questions, we divided the project into three parts: Text Analytics, Social Network Analytics, and Predictive Analytics.

In the first part, we did some Exploratory Data Analysis (EDA), and we found that 20.64% of all transactions are emoji only transactions. Food, People, and Transportation are the top three most popular emoji categories. We then created a dynamic spending profile for each user over her lifetime in Venmo in monthly intervals. For example, assume a user's first transaction was a pizza emoji. Then her user profile at 0 would be 100% food. Now, by the end of her first month in Venmo, she has transacted 4 times, 2 of them are food and 2 are activity related. Her spending profile in 1 month is 50% food and 50% activity. After having the spending profile for individual users, we then moved to compute the average and standard deviation of each spending category across all users. In general, we found that the trend of spending behaviors is growing with the time of usage, after the first two months, customer's spending behavior are stabilized, and it shows a growing trend. We could understand that the customers tend to engage more after they have some experience with the app.

In the second part, we focused on analyzing user's social network through three social network metrics across Her/His lifetime in Venmo. The first metric is Number of friends and number of friends of friends. (Friend definition: A user's friend is someone who has transacted with the user, either sending money to the user or receiving money from the

user). The second metric is clustering coefficient of a user's network. The clustering coefficient is a way to measure how closely connected individuals or entities are within a social network. It looks at how likely it is that people who are connected to the same person are also connected to each other, forming small, tight-knit groups. A clustering coefficient should be ranging from 0 to 1; a value of 0 means that the user's friends did not have any transactions among them. The third metric is page rank. Page rank is a method used to figure out which people or things in a network are the most important. Imagine you have a group of people, and each person is connected to others by lines, like a web. Page rank helps to determine who is the most influential person in that web.

In the third part, we focused on predicting in advance how many times your customers will transact. The main approach is to use linear regression with metrics that we have created earlier to predict the total number of transactions a user will have by the end of their first year in Venmo. We first use recency and frequency as independent variable. Recency refers to the last time a user was active, and frequency is how often a user uses Venmo in month. For example, if a user has used Venmo twice during his first month in Venmo with the second time being on day x, then her recency in month 1 is "30-x" and her frequency is 30/2 = 15. The second approach is to include user's spending behavior profile in addition to recency and frequency. The third approach is to regress social network metrics on Y (total number of transactions at lifetime point 12). Finally, the last approach is to regress social network metrics and the spending behavior profile on Y. By comparing the mean squared error (MSE) for all these approaches, we found that regress social network metrics on Y generate the best result. The most informative predictor is the page rank as it has the largest coefficient among all features, suggesting that the number of transactions could depend more on the authority or influence of a certain user than on the number of friends he/she has.

In conclusion, this project focused on Text Analytics, Social Network Analytics, and Predictive Analytics of Venmo's transaction data. We achieved the initial goal by having a comprehensive understanding of user's spending behavior profile, user's social network, and predict in advance how many times you customers will transact.